

ABSTRACT

A system for compressing text using variable length codes includes a memory device and an encoder. The memory device is configured to store a set of variable length codes for a plurality of languages for compression of a text, where the set of variable length codes includes variable code lengths based on language features. The encoder is coupled to the memory device and is configured to receive text in at least one of the plurality of languages, to generate a compressed text by assigning a code to each word in the text based on codes from the set of variable length codes that are associated with the at least one language to the text and to generate at least one header to be inserted in the compressed text. The header includes information regarding the location in the compressed text of a subsequent change in code length.